4択クイズを連想問題として解く

外池昌嗣[†]　佐藤理史[†]　宇津呂武仁[†]

解の選択は質問応答システムのコンポーネントの1つで、何らかの方法で得られた解候補の中から信頼できるものを選ぶものである。本研究では語と語の連想の強さに注目して解の選択を行う。本稿では、問題文から抽出されたキーワードと解候補の間の語彙的な関係の強さに基づく解の選択法を提案する。提案する解の選択法は2つのステップに分けられる。1つ目のステップでは、語の特徴と語彙的関係の強さを用いて、問題文から適切なキーワードを抽出する。一方、2つ目のステップでは、サーチエンジンのヒット数に基づいて、キーワードと解候補の間の関係の強さを測定する。実験の結果、提案手法で4択クイズ「クイズ$ミリオネア」の79%を解くことができた。

Answer Validation by Keyword Association

Masatsugu Tonoike ,[†] Satoshi Sato [†] and Takehito Utsuro[†]

Answer validation is a component of question answering system, which selects reliable answer from answer candidates extracted by certain methods. In this paper, we propose an approach of answer validation based on the strengths of lexical association between the keywords extracted from a question sentence and each answer candidate. The proposed answer validation process is decomposed into two steps: the first is to extract appropriate keywords from a question sentence using word features and the strength of lexical association, while the second is to estimate the strength of the association between the keywords and an answer candidate based on the hits of search engines. In the result of experimental evaluation, we show that a good proportion (79%) of a multiple-choice quiz "Who wants to be a millionaire" can be solved by the proposed method.

1. はじめに

自然言語で与えられた質問の答えを探す技術は質問応答と呼ばれ、近年注目を集めている。TREC QA Task[1)] やNTCIRのQAC[2)]など評価型ワークショップは、質問応答の研究を活発にさせた。質問応答のシステムは、大きく2つのステップに分けられる。一つは解候補を集めるステップで、もう一つはそれぞれの解の選択である。最初の、解候補を集めるステップは、これまでよく研究されてきた。このステップの標準的な手順では、まず「場所」、「人名」といった質問タイプを判定する。次に使用可能なドキュメント集合に対して、質問文から生成されるクエリーで検索を行い、解候補を含むドキュメントを収集する。そして、得られたドキュメントに対して固有表現抽出を行い、答えのタイプと合う語を解候補として抽出する。

本研究では、このような解候補抽出ではなく、質問応答の2つ目のステップである、解の選択に注目する。解の選択法として、これまでにいくつかの方法が提案されている。よく知られている方法の一つはテキストの深い理解に基づく。このような研究にはMoldovan[3)]らの研究がある。彼らの深い理解に基づく解の選択法ではまず、問題文と解候補を含むパラグラフを論理表現に変換する。そして、推論の操作によって両者が対応しているかを調べる。この方法の欠点の一つは、WordNetなどの語彙的知識と、推論ルールセットなどの世界知識という豊富な情報を必要とすることである。これに対して、本研究ではこれと異なったアプローチの解の選択法を提案する。その方法は、問題文中の重要語句と解候補の間の語彙的関係の強さの比較に基づいている。語彙的関係の強さを測定することは最も低いレベルの意味的操作と言えるが、本研究では、このような意味的操作が、質問応答における解の選択という豊富な知識を要求する自然言語処理タスクに対して、どの程度有効性かを調べる。本稿で後に示すように、問題の解候補として選択肢が与えられると、我々の語彙的関係に基づく方法で、ある種類の問題セットのかなりの問題を解くことができる。

† 京都大学大学院 情報学研究科
Graduate School of Informatics, Kyoto University

我々のキーワードアソシエーション (本稿では上記で導入した語彙的関係の概念を「キーワードアソシエーション」と呼ぶ) による解の選択法の枠組みにおいては、解の選択プロセスは 2 つのステップに分解できる。最初のステップでは、質問文から適切なキーワードを抽出する。そして、次のステップではキーワードと解候補の関係の強さを測定する。本稿で提案するキーワード選択法は、それぞれの語の特徴に基づいて人手で作成した少数のルールによって重要度を決める方法と、サーチエンジンのヒット数を利用して連想の強さの測定する方法の 2 つである。解候補の検証の二番目のステップでは、ウェブを知識源として用いて、抽出されたキーワードと解候補の間の関係の強さを測定する。キーワードと解候補の関係が強いほど同一ウェブページ上に共起することが多いという仮定に基づき、本稿ではこの関係の強さを、サーチエンジンのヒット数を利用したいくつかの尺度で測定し、それらの尺度の有効性を実験によって示す。

本研究では、質問応答全体のプロセスではなく解の選択の問題に焦点を当てるために、既存の 4 択クイズを研究の実験材料とする。本研究で題材として使う 4 択クイズは “Who wants to be a millionaire?”[☆] である。“Who wants to be a millionaire?” はイギリスに始まり、５０ヶ国以上の国でローカライズされて放映されている人気のあるテレビのクイズ番組である。今回はその中でも日本版の「クイズ＄ミリオネア」[☆☆4)] を用いる。実験の結果、このクイズの約 8 割の問題を、提案するキーワードアソシエーションによる解の選択法で解くことができる。

2 章では、キーワードアソシエーションを用いた質問応答の考え方を導入する。3 章では、問題文からキーワードを選ぶ方法を説明する。4 章では、 4 択クイズの答えを選ぶ方法を説明する。5 章では、提案手法を組み合わせる方法を説明する。6 章では、実験の結果を示す。

2. キーワードアソシエーションによる解の選択

2.1 キーワードアソシエーション

ここに 4 択クイズの例がある。

Q1: 「アメリカングラフィティ」の監督は誰？

a: ジョージ・ルーカス
b: スティーブン・スピルバーグ
c: フランシス・フォード・コッポラ
d: 黒澤明

いま、この問題の答えが分からず、ウェブのサーチエンジンを使って答えを見つけようとしているとしよう。もっとも単純な方法は、クエリー「アメリカングラフィティ」をサーチエンジンに入力し、得られたウェブページを調べるというものである。この戦略は、キーワード「アメリカングラフィティ」を含むウェブページ上に正解が現れることを仮定している。もう少し巧妙な方法は、キーワードと選択肢の両方を含むウェブページの数を使うというものである。この数は、「アメリカングラフィティ」と「ジョージ・ルーカス」の両方の語からなるクエリーを入力したときのサーチエンジンのヒット数で測定できる。この仮定に基づき、最大のヒット数を持つ選択肢が解であると決定する。上記の Q1 に対しては、この戦略は成功する。表 1 にそれぞれの選択肢のアンド検索のヒット数を掲載する。このとき、ヒット数を求めるのに goo[☆☆☆] を用いた。ここで、X はキーワードの集合、Y は選択肢を表すものとし、関数 $hits$ を以下のように定義する。

表 1 問題 Q1 におけるキーワードと選択肢のヒット数 (X:“アメリカングラフィティ”)

Y(選択肢)	hits(X and Y)
“ジョージ・ルーカス”	384
“スティーブン・スピルバーグ”	123
“フランシス・フォード・コッポラ”	100
“黒澤明”	82

$$hits(X) \equiv hits(x_1 \text{ AND } x_2 \text{ AND} \cdots \text{AND } x_n)$$

where

$$X = \{x_1, x_2, \ldots, x_n\}$$

正解である「ジョージ・ルーカス」のアンド検索のクエリーは、最大のヒット数を返す。

ここで、問題 Q1 はキーワードと選択肢の関係の強さに関する問題、つまり連想問題と見なすことができ、以下の形式に変換できる。

Q1′: 次のうち、「アメリカングラフィティ」と最も関係の深いものはどれ？

a: ジョージ・ルーカス
b: スティーブン・スピルバーグ
c: フランシス・フォード・コッポラ
d: 黒澤明

我々はこのキーワードと選択肢の関係のことを「キーワードアソシエーション」と呼ぶ.

2.2 どのようにキーワードを選ぶか

精度よく問題を解くためには、問題文から適切なキーワードを取り出すことが重要である。以下の問題を考えよう。

Q2: 映画「ロードオブザリング」の監督は誰？

a: イライジャ・ウッド
b: ビリー・ボイド
c: ピーター・ジャクソン
d: ケイト・ブランシェット

☆ Celador International 製作
☆☆ フジテレビ製作 http://www.fujiint.co.jp/quiz/home/index.html
☆☆☆ http://www.goo.ne.jp

表 2　問題 Q2 におけるキーワードと選択肢のヒット数 (X:“ロードオブザリング”, X':“ロードオブザリング” and “監督”)

Y(選択肢)	hits (X and Y)	hits (X' and Y)
“イライジャ・ウッド”	699	281
“ビリー・ボイド”	105	72
“ピーター・ジャクソン”	401	365
“ケイト・ブランシェット”	127	91

実際にヒット数を求めた結果を表 2 に示す。ここで、X は “ロードオブザリング”、 X' は “ロードオブザリング” and “監督” とする。 映画のタイトルである「ロードオブザリング」をキーワードとしたとき、アンド検索のヒット数が最大の選択肢は「イライジャ・ウッド」であり、正解の「ピーター・ジャクソン」ではない。しかし、「ロードオブザリング」と「監督」をキーワードとすれば、この問題は最大のヒット数を持つ選択肢を選ぶことによって解ける。したがって、適切なキーワードを選びだすメカニズムが必要となる。

2.3 フォワードアソシエーションとバックワードアソシエーション

キーワードと選択肢からなるクエリーを生成して、単純に最大のヒット数を持つ選択肢を選ぶ方法では、解けない問題もある。本節では、適切な答えを選ぶためのより緻密な尺度を導入する。以下の問題を見よ。

Q3: ピラミッドはどこにある？

a: カナダ
b: エジプト
c: 日本
d: 中国

この問題に対して、実際にヒット数を求めた結果を表 3 に示す。 この場合、キーワードと選択肢をクエリーとしたときのヒット数が最大の選択肢は「日本」であり、これは正解の「エジプト」ではない。なぜこの問題が解けないのであろうか。ここで選択肢のみをクエリーとしたときのヒット数に注目しよう。「日本」のみをクエリーとした場合のヒット数は、「エジプト」のみをクエリーとした場合のヒット数の約 24 倍である。この観察より、「ピラミッド」と選択肢をクエリーとしたときのヒット数は、選択肢のみのヒット数の影響を受けていると考えられる。それゆえ、正規化が必要となる。

上記の仮定に基づき、我々は佐藤らの提案する、2 つの語の関係の強さを評価するための 2 つの尺度[5] を導入する。X はキーワードの集合、Y は選択肢とする。本稿では、キーワードの集合 X と選択肢 Y からなるクエリーのヒット数を X のヒット数で正規化したものをフォワードアソシエーション $FA(X,Y)$ と呼ぶ。また、キーワードの集合 X と選択肢 Y からなるクエリーのヒット数を Y のヒット数で正規化したものをバックワードアソシエーション $BA(X,Y)$ と呼ぶ。

$$FA(X,Y) = hits(X \cup \{Y\})/hits(X)$$
$$BA(X,Y) = hits(X \cup \{Y\})/hits(\{Y\})$$

ここで、X が固定されていれば、FA は $hits(X \cup \{Y\})$ に比例する。

Q3 に戻ろう。この場合、最大の BA を持つ選択肢が正解である。ある問題は FA を参照することで解ける一方、BA を参照しなければ解けない問題もある。それゆえ、FA と BA のどちらを使うのかを決める仕組みを考案する必要がある。

2.4 ま と め

2.1 ∼ 2.3 節での観察より、連想による解の選択を実現するために、以下の 3 つの問題を解決する必要がある。

- どのようにして問題文から適切なキーワードを抽出するか？
- どのようにしてフォワードとバックワードの連想を考慮して正解を同定すればよいか？
- どれぐらいの問題を連想に基づく方法で解くことができるのか？

3. キーワード選択

本章では問題文から適切なキーワードを選ぶ 2 つの手法について述べる。一つは語の特徴を用いる手法で、もう一つはサーチエンジンのヒット数を用いる手法である。

まず、日本語形態素解析システム Juman[☆] と日本語構文解析システム KNP[☆☆] を用いて、問題文からすべての名詞を抽出する。このとき、名詞の列が複合語を構成するとき、最も長い複合語のみを抽出し、その構成要素である名詞は抽出しない。このようにして抽出された名詞や複合語の集合を N とする。そして、キーワードはその中から選ばれる。以下に説明する手法では、サーチエンジン goo を使用して、ヒット数を求める。

3.1 語の特徴に基づくキーワード選択

語の特徴に基づくキーワード選択手法では、キーワードは以下の手続きによって選ばれる。

(1) 問題文が n 個の「 」で括られた引用を含むならば、それらの n 個の文字列がキーワードとして選ばれる。
(2) それ以外の場合:
  2-1. 表 4 に示す語の重み付けルールにしたがって、キーワード候補の集合 N のそれぞれの要素に重みが付与される。
  2-2. 最大及び 2 番目に大きな重みを持つキーワー

☆ http://www.kc.t.u-tokyo.ac.jp/nl-resource/juman.html
☆☆ http://www.kc.t.u-tokyo.ac.jp/nl-resource/knp.html

表 3　問題 Q3 におけるキーワードと選択肢のヒット数

X(キーワード)	hits(X)			
ピラミッド	63,600			
Y(選択肢)	hits(Y)	hits(Y and X)	FA(X, Y)	BA(X, Y)
カナダ	321,000	3,300	0.0159	0.0103
エジプト	128,000	14,300	0.225	0.112
日本	3,050,000	27,800	0.437	0.00911
中国	2,580,000	10,300	0.162	0.00399

表 4　語の重み付けルール

条件	重み
n 番目の文節	1 + 0.01 * n
ストップワード	0
「 」中の語句	3
人名	3
サ変名詞 + サ変動詞	0.5
関係を表す語	2
カタカナ (重要でないものを除く)	2
「賞」で終わる	2
「時代」で終わる	0.5
国名	0.5
数字	3
ヒット数が 1000000 より大 and 1 文字	0.9
キーワードの後ろに「は」が存在 and	
「者」または「家」で終わる	0.1
ヒット数が 100000 より大	0.2
ヒット数が 10000 より小	1.1
文字数が 1	0.2
文字数が 2	0.25
文字数が 3	0.5
文字数が 4	1.1
文字数が 5 以上	1.2

ド候補を選択する。

2-3. (i) これら 2 つのキーワードの候補語の AND 検索のヒット数が 15 以上であれば、両方をキーワードとして選択する。

(ii) そうでなければ、重み 1 位のキーワードのみを選択する。

k を選ばれたキーワードの集合 ($k \subseteq N$)、c を選択肢として、$c_1^{FA}(k)$ を最大の $FA(k,c)$ を持つ選択肢、$c_1^{BA}(k)$ を最大の $BA(k,c)$ を持つ選択肢と定義する。

$$c_1^{FA}(k) = \underset{c}{\mathbf{argmax}}\, FA(k,c)$$
$$c_1^{BA}(k) = \underset{c}{\mathbf{argmax}}\, BA(k,c)$$

ここで、$c_1^{FA}(k)$ か $c_1^{BA}(k)$ のどちらかが正解であれば、このとき選択されたキーワードは正解であるとすると、6.1 節で導入される分析用セットに対する正解率は、84.5%であった。

3.2 サーチエンジンのヒット数に基づくキーワード選択

3.2.1 基本的な手法

まず、サーチエンジンのヒット数に基づく単純なキーワード選択法をいくつか導入する。ここで、2^N は N の巾乗を示すものとすると、キーワードの集合 k は 2^N の要素である ($k \in 2^N$)。$\hat{k}$ は選ばれたキーワード集合、$\hat{c}$ は選ばれた答えを示す。

最初の手法は、以下のように最大のヒット数を持つ $<\hat{k}, \hat{c}>$ のペアを選ぶものである。

$$<\hat{k}, \hat{c}> = \underset{c,\ k \subseteq 2^N}{\mathbf{argmax}}\, hits(k \cup \{c\})$$

分析用セットに対して、この手法で選ばれた選択肢の正解率を計算すると 35.7% であった。

同様に、最大の FA または BA を選ぶ別の手法を以下に示す。

$$<\hat{k}, \hat{c}> = \underset{c,\ k \subseteq 2^N}{\mathbf{argmax}}\, FA(k \cup \{c\})$$
$$<\hat{k}, \hat{c}> = \underset{c,\ k \subseteq 2^N}{\mathbf{argmax}}\, BA(k \cup \{c\})$$

正解率はそれぞれ、71.3%と 36.1%であった。これまでに示した 3 つの手法の結果を表 5 にまとめる。

次に、$c_1^{FA}(k)$ と $c_1^{BA}(k)$ が一致していれば、そのとき選ばれたキーワード集合と答えは信頼できるのではないかという仮説に基づいて以下の調査を行った。ここで、$\hat{K}$ を選ばれたキーワード集合の集合とする。

$$\hat{K} = \{k | \underset{c}{\mathbf{argmax}}\, FA(k,c) = \underset{c}{\mathbf{argmax}}\, BA(k,c)\}$$

分析用セットに対して、この方法でどの程度キーワード集合を選べるのかを調べた結果を表 6 に示す。この結果、キーワード集合が 1 つに絞れたのは 70 問 (7.9%) であった。キーワード集合が 1 つに絞れなかった 700 問 (78.8%) のうち、すべての答えが一致したのは 415 問で、全体の

表 5　ヒット数/FA/BA が最大のキーワードを選ぶ手法

手法	正解率
ヒット数	35.7%
FA	71.3%
BA	36.1%

表 6　FA と BA が一致

条件	個数
$\lvert\hat{K}\rvert = 1$	70 (7.9%)
$\lvert\hat{K}\rvert \geq 2$	700 (78.8%)
$\lvert\hat{K}\rvert = 0$	118 (13.3%)
合計	888 (100%)

表 7　FA 最大と BA 最大が一致

条件	個数
$\|\hat{K}\| = 1$	13 (1.5%)
$\|\hat{K}\| = 0$	875 (98.5%)
合計	888 (100%)

46.7%(415/888) であった。キーワード集合が 1 つに絞れた場合 (70 問) と、キーワード集合は 1 つに絞れなかったがすべての答えが一致した場合 (415 問) は、答えを決めることができる。答を決められた 485 問 (全体の 54.6%) の精度は 82.1%(398/485) であった。

さらに条件を厳しくして、$FA1$ 位の集合キーワードと $BA1$ 位のキーワード集合が一致する場合のみ、それを採用するという方法も調査した。この方法を以下に式で示す。

$$\exists c, \hat{K} = \{k | \operatorname*{\mathbf{argmax}}_{k \subseteq N} FA(k,c) = \operatorname*{\mathbf{argmax}}_{k \subseteq N} BA(k,c)\}$$

この方法でどの程度キーワード集合を選べるのかを調べた結果を表 7 に示す。しかしながら、わずかに 13 問しか一致しなかった。

このような手法では、キーワード集合を選ぶ条件が厳しすぎるため、適用できる問題が限られてしまう。そこで、次節ではここで説明した手法よりも条件の緩い手法を検討する。

3.2.2　キーワードアソシエーションの比

次に、我々は FA や BA といった尺度の 1 位と 2 位の比を用いる、より緻密な方法を導入する。この手法の根底にある仮説は、それらの比が大きくなれば、選ばれた FA もしくは BA が最大の選択肢の信頼性が増すということである。ここではまず、それぞれ FA の比を用いる手法と、BA の比を用いる手法の、 2 つの手法を導入する。

FA の 比

FA の比とは、あるキーワード集合に対して、2 番目に大きな FA を持つ選択肢の FA を、最大の FA を持つ選択肢の FA で割った値である。FA の比を用いる手法では、この比を用いてキーワード集合を評価する。実際の手順ではまず、このような FA の比をすべてのキーワード集合に対して計算し、最大の比の値を持つキーワード集合を選ぶ。そして、このときの FA 最大の選択肢を解とする。選ばれたキーワード集合を $\hat{k}$、選ばれた選択肢を $\hat{c}$ とすると、この方法は以下のように定式化できる。

$$\begin{aligned}\hat{k} &= \operatorname*{\mathbf{argmin}}_{k \subseteq 2^N} \frac{FA(k, c_2^{FA}(k))}{FA(k, c_1^{FA}(k))} \\ \hat{c} &= c_1^{FA}(\hat{k}) \\ c_2^{FA}(k) &= \operatorname*{\mathbf{arg\text{-}secondmax}}_{c} FA(k,c)\end{aligned}$$

ただし、$\mathbf{arg\text{-}secondmax}_c$ は値が 2 番目に大きくなるときの、c を選ぶ関数である。

表 8　キーワードアソシエーションの比の評価 (精度/カバレージ)(%)

		1 位、2 位	
		FA	BA
比	FA	63.1/100	70.6/95.0
	BA	**75.8/93.2**	67.6/100

同様に BA の比に基づく手法を以下に示す。

BA の 比

$$\begin{aligned}\hat{k} &= \operatorname*{\mathbf{argmin}}_{k} \frac{BA(k, c_2^{BA}(k))}{BA(k, c_1^{BA}(k))} \\ \hat{c} &= c_1^{BA}(\hat{k}) \\ c_2^{BA}(k) &= \operatorname*{\mathbf{arg\text{-}secondmax}}_{c} BA(k,c)\end{aligned}$$

次は、FA の比や BA の比に基づく方法と違い、FA と BA の両方を考慮する方法を述べる。これら 2 つの手法の仮説は、FA と BA が指し示す選択肢が同じであれば、FA と BA による解の決定は信頼できると見なすというものである。

FA が 1 位と 2 位の選択肢の BA の比

それぞれのキーワードに対して、FA が 1 位と 2 位の選択肢を選び、それらの選択肢の BA の比を計算する。そして、この比の値が最も大きかったキーワードを選び、$\hat{k}$ とする。そして、$c_1^{FA}(\hat{k})$ をこの手法の選んだ解とする。

$$\begin{aligned}\hat{k} &= \operatorname*{\mathbf{argmin}}_{k} \frac{BA(k, c_2^{FA}(k))}{BA(k, c_1^{FA}(k))} \\ \hat{c} &= c_1^{FA}(\hat{k})\end{aligned}$$

同様に、BA が 1 位と 2 位の選択肢の FA の比を用いる手法を以下に示す。

BA が 1 位と 2 位の選択肢の FA の比

$$\begin{aligned}\hat{k} &= \operatorname*{\mathbf{argmin}}_{k} \frac{FA(k, c_2^{BA}(k))}{FA(k, c_1^{BA}(k))} \\ \hat{c} &= c_1^{BA}(\hat{k})\end{aligned}$$

これら 4 つの手法のカバレージと精度を表 8 に示す。カバレージは比が 1 以下の問題の割合によって測定する⋆。精度は、選ばれた選択肢 $\hat{c}$ が正解である問題数を、上記のカバーされている問題数で割った値によって測定される。最もよい精度を示したのは、最大の FA を持つ選択肢と 2 番目に大きい FA を持つ選択肢の BA の比を用いる手法であった。以下の章では、この比を「キーワードアソシエーションの比」として用いる。比の値の範囲とカバレージ/精度の関係を表 9 に示す。比が 0.25 以下であるとき、全体の約 60%の問題が 90%近くの精度で解けた。この 0.25 という閾値は、5 章でキーワードアソシ

表 9　キーワードアソシエーションの比の評価: FA が 1 位と 2 位の選択肢の BA の比

比	# of questions	
	カバレージ	精度
0	18.9% (163/888)	89.6% (146/163)
≤ 0.01	21.5% (191/888)	89.5% (171/191)
≤ 0.1	40.5% (360/888)	87.5% (315/360)
$\mathbf{\leq 0.25}$	**60.4% (536/888)**	**86.9% (466/536)**
≤ 0.5	78.0% (693/888)	81.6% (566/693)
≤ 0.75	87.2% (774/888)	78.4% (607/774)
≤ 1	93.2% (828/888)	75.8% (628/828)

エーションの比と語の重みを組み合わせるときに用いる。

4. 解 の 選 択

本章では、フォワード及びバックワードアソシエーションを考慮して正解を同定する方法を説明する。キーワードを決定した後は、サーチエンジンを利用して、以下の値を求める。

- キーワード集合 X のヒット数: $hits(X)$
- 選択肢 Y のヒット数: $hits(\{Y\})$
- AND 検索のヒット数: $hits(X \cup \{Y\})$

そして、それぞれの選択肢に対して、FA と BA が計算される。3 章で導入したように、$c_1^{FA}(k)$ はあるキーワード集合 k が与えられたときに最大の FA を持つ選択肢を表し、$c_1^{BA}(k)$ は同様に k が与えられたときに最大の BA を持つ選択肢を表す。ここですべきことは、$c_1^{FA}(k)$ か $c_1^{BA}(k)$ のどちらが正解かを推測することである。

分析用セットにおいてサーチエンジンのヒット数を調査し、以下に示す $c_1^{FA}(k)$ と $c_1^{BA}(k)$ を選択するルールを人手で作成した。

(1) if $c_1^{FA}(k) = c_1^{BA}(k)$ then $c_1^{FA}(k)$

(2) else if $\frac{FA(k,c_1^{BA}(k))}{FA(k,c_1^{FA}(k))} \geq 0.8$ then $c_1^{BA}(k)$

(3) else if $\frac{FA(k,c_1^{BA}(k))}{FA(k,c_1^{FA}(k))} \leq 0.2$ then $c_1^{FA}(k)$

(4) else if $\frac{BA(k,c_1^{FA}(k))}{BA(k,c_1^{BA}(k))} \geq 0.53$ then $c_1^{FA}(k)$

(5) else if $hits(k) \geq 1300$ then $c_1^{BA}(k)$

(6) else if $\frac{FA(k,c_1^{BA}(k))}{FA(k,c_1^{FA}(k))} \geq 0.6$ then $c_1^{BA}(k)$

(7) else $c_1^{FA}(k)$

表 10 は、分析用セットに対して、この解選択法を適用したときの精度を示す。ここで、キーワードは 3.1 節の語の重みに基づく手法で選んだものを用いた。この表では、上記の解選択ルールの結果に加えて、単純に FA または BA が最大の選択肢を選ぶ、ベースラインとなる手法の結果も示した。この結果より、ここで述べた解選択ルールはベースラインの手法の精度を大きく上回ること

☆ この比は FA と BA の両方を考慮するので、比が 1 より大であることは FA と BA が指し示す選択肢が異なることを意味する。

表 10　解選択の精度 (語の重みに基づくキーワード選択との組み合わせ)

FA 最大	70.8%
BA 最大	67.6%
解選択ルール	77.3%

表 11　それぞれの解選択ルールの評価 (語の重みに基づくキーワード選択を用いる)

ルール	答え	精度
1	$c_1^{FA}(k) = c_1^{BA}(k)$	88.5% (479/541)
$2 \sim 6$	-	60.3% (207/343)
合計	-	77.6% (686/884)
2	$c_1^{BA}(k)$	65.3% (32/49)
3	$c_1^{FA}(k)$	61.8% (68/110)
4	$c_1^{FA}(k)$	53.6% (37/69)
5	$c_1^{BA}(k)$	60.3% (35/58)
6	$c_1^{BA}(k)$	66.7% (12/18)
7	$c_1^{FA}(k)$	59.0% (23/39)

が分かった。

解選択ルールのそれぞれに対して、表 11 にその精度を示す。分析用セットにおいて☆、$c_1^{FA}(k)$ と $c_1^{BA}(k)$ が一致する問題は 541 問 (約 60%) あり、選ばれた選択肢の 88.5%が正解であった。つまり、半数以上の問題の $c_1^{FA}(k)$ と $c_1^{BA}(k)$ が一致し、このうち 9 割以上は正解であった。したがって、この $c_1^{FA}(k)$ と $c_1^{BA}(k)$ が一致すれば、選ばれた解は信頼できる。

5. キーワード選択と解選択の統合手法

3.2.2 節で示したキーワードアソシエーションの比を用いる手法では、比が 0.25 以下であれば高い精度で問題を解くことができた。一方、3.1 節で示した語の重みに基づくキーワード選択法と 4 章で示した解選択ルールを組み合せた手法は、分析用セットの問題すべてを解いたときには、キーワードアソシエーションの比を用いる手法よりも精度がよかった。そこで、両者の長所を生かすために、以下に示す統合手法を実装した。

(1) 比が 0.25 以下のとき:
FA が 1 位と 2 位の選択肢の BA の比によって選ばれたキーワードの集合を選択し、最大の BA を持つ選択肢を選ぶ。

(2) それ以外の場合:
語の重みに基づく方法で選ばれたキーワードの集合を使い、4 章の手順によって、解選択を行う。

6. 実　　験

6.1 データセット

本研究では、4 択クイズとして、トミー社から発売されているカードゲーム版の「クイズ $ ミリオネア」の問題を用いた。カードゲーム版の問題は、1960 問あり、そ

☆ アンド検索のヒット数 $hits(X \cup \{Y\})$ が 0 であった 4 問は除く

表 12　最終的な実験結果 (精度/カバレージ)(%)

手法	分析用	評価用
K.A.R. ($r \leq 1$)	75.8/93.2	74.6/93.6
語の重み ＋ 解選択ルール	77.3/100	73.4/100
統合手法	78.6/100	75.9/100
K.A.R. ($r \leq 0.25$)	86.9/60.4	86.0/61.5
語の重み ($r > 0.25$) ＋ 解選択ルール	65.9/39.6	59.9/38.5

K.A.R.: キーワードアソシエーションの比

れらは、賞金額別に A(￥10,000) ~ O (￥10,000,000) の 15 クラスからなる。それぞれの問題は解候補として 4 つの選択肢を持つ。

ここでは、それぞれのクラスを 2 つに分割し、半分 (980 問) を分析用の問題セットとして、残り半分を評価用の問題セットとして使用する。ただし、比較表現を含む問題 (例えば、「次のうち、最も州の数が多い国はどれ？」) と否定の表現を含む問題 (例えば、「次のうち、フランスの国旗に使われていない色はどれ？」) は連想で説くのにはふさわしくないので除外した。その結果、分析用セットの問題数は 888 問に、評価用セットの問題数は 906 問になった。

6.2 結　　果

分析用セットと評価用セットに対して、5 章で示したキーワード選択と解選択の統合手法の評価の結果を表 12 に示す。この表では、3.2.2 節で示したキーワードアソシエーションの比を用いる手法と、3.1 節で示した語の重みに基づく手法及び、4 章で示した解選択を組み合わせた手法という、2 つのベースラインの性能も示している。5 章で示した統合手法は、それらのベースラインの性能を上回った。結果として、約 79%(分析用セット) あるいは 76%(評価用セット) の問題を、5 章で示した統合手法により解くことができた。

2 つの問題セットを比較すると、語の重み付けルールと解選択ルールを組み合わせた手法では、評価用セットにおいて精度が約 4%低下した。これは、語の重み付けルールと解選択ルールが分析用セットに過適応していることを示している。一方、キーワードアソシエーションの比を用いる手法においては、2 つの問題セットの間で精度の差はそれほどなかった。これは、キーワードアソシエーションの比を用いる手法はデータに依存せず有効であることを示している。

7. 関 連 研 究

本章では、ウェブを利用した関連研究を取り上げ、本研究との違いを述べる。

Cody ら[6] はウェブを用いた最初の質問応答システムを提案した。まず、google を使用して質問文に関係のない文書を集め、そこから解候補を抽出する。そして、キーワードの近くに現れる解候補の出現頻度に基づいて解を選択する。

Brill ら[7] によって提案された方法では、解候補はサーチエンジンの検索結果のサマリーから抽出する。そして、TREC QA タスクのドキュメントコレクションから関係する文書を探すことによって、それぞれの解候補の検証を行う。Cody らの手法、Brill らの手法共に、サーチエンジンのヒット数は利用していない。

Magnini ら[8] はサーチエンジンのヒット数を使う解の選択法を提案した。彼らは AltaVista の OR と NEAR の演算子を使ってサーチエンジンのクエリーを構成する。Magnini らの手法と我々の手法の主要な違いは、キーワードの選択法にある。Magnini らの手法において、キーワード集合の初期候補は質問文中にあるすべての内容語を含む。キーワード集合のヒット数がある閾値以下であれば、人手で作成した少数の規則に従い、キーワード集合から最も重要でない語を除く。この手続きをキーワードのヒット数が閾値を超えるまで繰り返す。一方、我々の方法では、キーワードと解候補の関係の強さが最大になるようなキーワードを選ぶ。Magnini らの方法に比べると、我々のキーワード抽出法は問題文と解候補の両方を考慮する点が異なっている。Magnini らの実験では、解候補にスコアをつける 3 つの尺度のうち、我々のフォワードアソシエーションに近い “Corrected Conditional Probability” という尺度が最もよい性能であった。我々もこの尺度を実装したところ、5 章で示した統合手法よりも約 5%低い精度であった。

8. 結論と今後の課題

本稿では、問題文から抽出されたキーワードと解候補の間の語彙的な関係の強さに基づく解の選択法を提案した。提案した解の選択法は 2 つのステップに分けられる。1 つ目のステップでは、語の特徴と語彙的関係の強さを用いて、問題文から適切なキーワードを抽出する。一方、2 つ目のステップでは、サーチエンジンのヒット数に基づいて、キーワードと解候補の間の関係の強さを測定する。実験の結果、提案手法で 4 択クイズ「クイズ $ ミリオネア」の 79%を解くことができた。

今後の課題として現在、問題文の構文構造が、問題文から適切なキーワードを選ぶのに有用かどうかの調査を行っている。また、本稿で提案したキーワード選択法を、質問応答のプロセス全体における解候補収集などの局面において利用する手法を設計する予定である。

参 考 文 献

1) Voorhees, E. M.: Overview of the TREC 2002 Question Answering Track, *Proc. TREC 2002* (2003).
2) Fukumoto, J., Kato, T. and Masui, F.: Question Answering Challenge (QAC-1) An Evaluation of

Question Answering Task at NTCIR Workshop 3, *Proc. 3rd NTCIR Workshop* (2003).
3) Moldovan, D., Harabagiu, S., Girju, R., Morarescu, P. and Lacatusu, F.: LCC Tools for Question Answering, *Proc. TREC 2002* (2003).
4) クイズ$ミリオネア (編): 完全攻略！クイズ$ミリオネア, フジテレビ出版 (2002).
5) Sato, S. and Sasaki, Y.: Automatic Collection of Related Terms from the Web, *Proc. 41st ACL*, pp. 121–124 (2003).
6) Kwok, C. C. T., Etzioni, O. and Weld, D. S.: Scaling Question Answering to the Web, *Proc. the 10th WWW Conf.*, pp. 150–161 (2001).
7) Brill, E., Lin, J., Banko, M., Dumais, S. and Ng, A.: Data-Intensive Question Answering, *Proc. TREC 2001* (2002).
8) Magnini, B., Negri, M., Prevete, R. and Tanev, H.: Is It the Right Answer? Exploiting Web Redundancy for Answer Validation, *Proc. 40th ACL*, pp. 425–432 (2002).
9) 外池昌嗣, 佐藤理史: ウェブを用いて４択クイズを解く, 言語処理学会第 9 回年次大会発表論文集, pp. 641–644 (2003).
10) 外池昌嗣, 佐藤理史, 宇津呂武仁: ４択クイズを連想問題として解く, 言語処理学会第 10 回年次大会発表論文集, pp. 301–304 (2004).